

Outage Efficient Strategies for Network MIMO with Partial CSIT

Mari Kobayashi¹, Sheng Yang¹, Mérouane Debbah¹, Jean-Claude Belfiore²

¹ SUPELEC, Gif-sur-Yvette, 91192, France

² Télécom ParisTech, 75013, Paris, France

Abstract

We consider a multi-cell MIMO downlink (network MIMO) where B base-stations (BS) with M antennas connected to a central station (CS) serve K single-antenna user terminals (UT). Although many works have shown the potential benefits of network MIMO, the conclusion critically depends on the underlying assumptions such as channel state information at transmitters (CSIT) and backhaul links. In this paper, by focusing on the impact of partial CSIT, we propose an outage-efficient strategy. Namely, with side information of all UT's messages and local CSIT, each BS applies zero-forcing (ZF) beamforming in a distributed manner. For a small number of UTs ($K \leq M$), the ZF beamforming creates K parallel MISO channels. Based on the statistical knowledge of these parallel channels, the CS performs a robust power allocation that simultaneously minimizes the outage probability of all UTs and achieves a diversity gain of $B(M - K + 1)$ per UT. With a large number of UTs ($K \geq M$), we propose a so-called distributed diversity scheduling (DDS) scheme to select a subset of \tilde{K} UTs with limited backhaul communication. It is proved that DDS achieves a diversity gain of $B \frac{K}{\tilde{K}} (M - \tilde{K} + 1)$, which scales optimally with the number of cooperative BSs B as well as UTs. Numerical results confirm that even under realistic assumptions such as partial CSIT and limited backhaul communications, network MIMO can offer high data rates with a sufficient reliability to individual UTs.

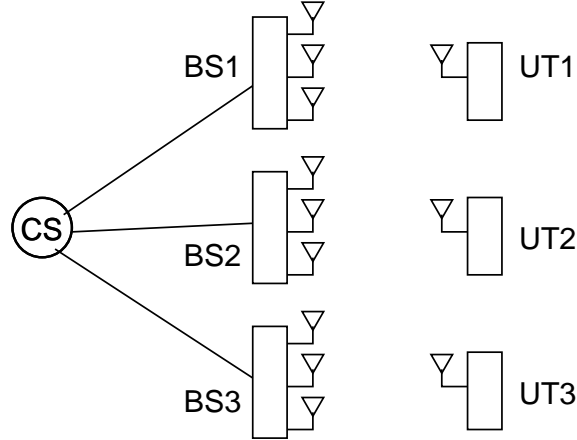


Fig. 1. A multicell downlink with $B = 3$ BSs and $K = 3$ UTs.

I. INTRODUCTION

Recently, network MIMO schemes, where neighboring Base-Stations (BSs) are connected to form an antenna array, have been proposed as a means to drastically increase the downlink capacity and solve the interference management problem of cellular systems [1]. Inspired by this result, we consider the multi-cell MIMO downlink where B BSs with M antennas connected to a Central Station (CS) communicate simultaneously with K User Terminals (UTs) with a single antenna each. Fig.1 illustrates an example of a multi-cell downlink system for $B = K = 3$ and $M = 3$. The channel at hand is modeled by the Multi-Input Single-Output (MISO) interference channel, defined by

$$y_k[t] = \sum_{i=1}^B \mathbf{h}_{ik}^H \mathbf{x}_i[t] + n_k[t] \quad (1)$$

for $t = 1, \dots, T$, where $y_k[t]$ is the channel output at UT k , \mathbf{h}_{ik}^H denotes the channel vector from BS i to UT k , $n_k[t] \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ is the Additive White Gaussian Noise (AWGN), and $\mathbf{x}_i \in \mathbb{C}^{M \times 1}$ denotes the input vector transmitted by BS i subject to the average power constraint P_i .

If the CS or equivalently all the BSs have perfect *Channel State Information at the Transmitter* (CSIT) and share the messages of all UTs, the channel at hand falls down into a classical $BM \times K$ MIMO broadcast channel with per-BS power constraints. In this case, the optimal strategy to maximize the multi-cell throughput is joint dirty-paper coding [1], [2], [3]. In order to capture the essential features of the multi-cell systems while enabling the analysis tractable, the Wyner model [4] has been widely considered in the literature. In [5], the authors provide a survey on the information theoretic results on

the multi-cell systems under the Wyner model for the Gaussian and fading channels with a single-antenna BS ($M = 1$). These include the per-cell downlink capacity based on the circular Wyner model [3], [6], [7] and the corresponding analysis in the different asymptotic regime such as high SNR, a large number of BSs, UTs [3], [6]. If each BS is equipped with multiple antennas ($M > 1$), the multi-cell downlink capacity can be naturally enhanced by exploiting the spatial degrees of freedom (see for example [8], [9], [10]). For a small B , the MIMO multi-cell downlink channel is also referred to the MIMO interference channel or MIMO-X channel under various message sharing assumptions (see [8], [11] and references therein). In these contributions, the sum degrees of freedom has been extensively studied.

Unfortunately, the global joint processing at the CS is difficult (if not impossible) in practice. This calls for practical network MIMO schemes which build on distributed processing at each BS by explicitly taking into account realistic aspects. A large number of recent contributions have been focused on practical designs by considering the following main limitations in network MIMO (see e.g. [12], [13], [14], [15], [16], [17], [18] and references therein). First, a substantial amount of resource needs to be dedicated for the CS to obtain accurate CSI. In particular, this overhead increases significantly with the number of cooperative BSs, which in turn leaves few resource for the data transmission within a fixed coherence block. In [12], such tradeoff between the benefits of network MIMO and the overhead in channel estimation has been studied for the case of the uplink with $M = 1$ and $B = K$. Second, the backbone links between the BSs and the CS are typically imperfect. They might be the capacity-limited [13], [14], [15], erroneous, or delayed [16]. This backhaul imperfectness will prevent the BSs from fully sharing the side information on the messages or CSI. Similar effects may occur when only adjacent BSs are connected to each other and exchange side information [17], [18].

Our contribution is no exception. We aim to design a practical scheme which ensures high data rates with sufficient reliability to individual UTs under partial CSIT. The objective at hand is relevant to most of the current/next wireless standards [19], [20]. To this end, we assume that each BS i has local CSIT while the CS has only statistical CSIT. Notice that the last assumption is reasonable because the CS needs to track the downlink channels at a rate much slower than the coherence time. We consider that the CS generates the messages destined to all K UTs and sends them to the B BS over the backbone link so that each BS locally encodes the messages and transmits. In order to concentrate on the impact of partial CSIT, we do not take into account other practical limitations. In particular, the underlying

backbone links are assumed to be perfect such that B BSs can fully share the messages¹. First, we address the case where the number of UTs is smaller than the number of transmit antennas, i.e., $K \leq M$. In this case, each BS applies ZF beamforming in a distributed manner, which creates K parallel MISO channels. Based on the statistical knowledge of these parallel channels, the CS performs a robust power allocation that simultaneously minimizes the outage probability of K UTs and achieves a diversity gain of $B(M - K + 1)$ per UT. Next, we consider a more relevant case of a large number of UTs ($K > M$). In this case, we propose a simple scheduling scheme called distributed diversity scheduling (DDS) which efficiently chooses a set of $\tilde{K} \leq M$ UTs while limiting the amount of the backhaul communication. More precisely, each BS i with local CSIT knowledge chooses its best set of UTs over the predefined partition and reports the corresponding index and value to the CS. The CS then decides and informs the selected set to all BSs. Finally, the selected UTs are served exactly in the same manner as the previous case of $K \leq M$. It is proved that DDS offers a diversity gain of $B \frac{K}{\tilde{K}} (M - \tilde{K} + 1)$ to each UT and moreover this gain scales optimally in B, K , and M , respectively. We remark that a similar CSIT assumption (local CSIT at BSs and statistical CSIT at the CS) has been also considered in [21]. However, the objective of this contribution is maximizing the multi-cell throughput rather than minimizing the outage probability as we address here.

Numerical results validate the analysis in terms of diversity gain and show that our proposed distributed ZF beamforming significantly outperforms the non-cooperative scheme. Namely, the outage performance of our proposed scheme improves with the number of cooperative BS, transmit antennas, and UTs. It is also shown in a simple one-dimensional topology that the performance gain can be emphasized as the path-loss between neighboring cells decreases. The main finding of this paper is that even in a realistic scenario with partial CSIT network MIMO can be beneficial by providing high data rate with a sufficient reliability to individual UTs. Such merit of network MIMO has been somehow overlooked in most of existing works assuming perfect CSIT.

The rest of the paper is organized as follows. The following section describes the system model. In Section III we present the power allocation policies that minimize the outage probability both under perfect and statistical CSI at the CS. We consider the relevant case of $K \geq M$ in Section IV where we propose a simple user scheduling algorithm and provide its diversity analysis. Section V shows some numerical results and Section VI concludes the paper.

¹If some backbone links are in failure, the corresponding BSs fail to encode the messages to K users. This will reduce the effective number of BSs.

Notations: Throughout the paper, we use boldface lower case letters to denote vectors, boldface capital letters to denote matrices. $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ respectively denote the complex conjugation, matrix transposition, and Hermitian transposition operations. \mathbf{I}_n and $\mathbf{0}_{n \times m}$ represent the $n \times n$ identity matrix and $n \times m$ zero matrix. The determinant, rank, trace, and Frobenius norm of a matrix \mathbf{A} are denoted by $|\mathbf{A}|$, $\text{rank}(\mathbf{A})$, $\text{tr}(\mathbf{A})$, and $\|\mathbf{A}\|_F^2$, respectively. The dot-equality stands for the near-zero equality, that is, $f(\epsilon) \doteq \epsilon^n$ means $\lim_{\epsilon \rightarrow 0} \frac{\log f(\epsilon)}{\log \epsilon} = n$. We let χ_m^2 denote the chi-square distribution with m degrees of freedom.

II. DISTRIBUTED ZERO-FORCING BEAMFORMING

In order to split the processing at the CS and the BSs, it is reasonable to assume that the CS generates the messages and performs the resource allocation whereas each BS encodes and then transmits the symbols in a distributed fashion. More precisely, we assume that the CS broadcasts K messages intended to K UTs which enable the BSs to cooperate in transmission. Then, each BS i encodes these messages into KT symbols $\{s_{ik}[t]\}$ for $k = 1, \dots, K$ and $t = 1, \dots, T$ by some capacity-achieving space-time coding (with a sufficiently large T channel uses). Under this setting, this section focuses on a distributed precoder design at each BS which requires only local CSIT $\{\mathbf{h}_{ik}\}_{k=1}^K$ for any i .

In the multi-cell downlink channel (1), we model the vector \mathbf{h}_{ik} of channel coefficients from BS i to UT k Gaussian distributed $\sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_{ik}\mathbf{I}_M)$ where the variance σ_{ik} captures the path-loss of the corresponding link assuming that the UTs are arbitrary distributed. Furthermore, $\{\mathbf{h}_{ik}\}$ are assumed to be i.i.d. over any pair i, k . We start with the definition of zero-forcing beamforming vectors as well as a useful lemma on the resulting channel statistics.

Definition 1 (Zero-forcing beamforming vectors): For a channel matrix $\mathbf{H} \in \mathbb{C}^{K \times M}$ with $K \leq M$ linearly independent row vectors $\mathbf{h}_1^H, \dots, \mathbf{h}_K^H$, there exists a zero-forcing beamforming matrix $\mathbf{G} = [\mathbf{g}_1, \dots, \mathbf{g}_K] \in \mathbb{C}^{M \times K}$, composed by K column vectors $\mathbf{g}_1, \dots, \mathbf{g}_K$, which is defined as

$$\mathbf{G} = \mathbf{H}^+ \text{diag}\{a_k\} \quad (2)$$

where $\mathbf{H}^+ = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1}$ and $\text{diag}\{a_k\}$ is a diagonal matrix that normalizes the norm of the columns of \mathbf{H}^+ such that $\|\mathbf{g}_k\| = 1$ for any k .

Lemma 1: If $\mathbf{H} \in \mathbb{C}^{K \times M}$ has i.i.d. entries where each row $\mathbf{h}_k^H \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_k \mathbf{I})$, then $|a_k|^2$ as defined in (2) is $\chi_{2(M-K+1)}^2$ distributed with mean $\mathbb{E}[|a_k|^2] = \sigma_k$, $\forall k$.

We consider a simple ZF beamforming which enables each UT to achieve a multiplexing gain of one. At each channel use t , BS i applies ZF beamforming to transmit K symbols in a distributed manner.

Due to the symmetry over all channel uses, we ignore the index t hereafter. We form the transmit vector of BS i at any channel use as

$$\mathbf{x}_i = \mathbf{G}_i \mathbf{u}_i = \sum_{k=1}^K \sqrt{p_{ik}} \mathbf{g}_{ik} s_{ik} \quad (3)$$

where we let $\mathbf{G}_i = [\mathbf{g}_{i1}, \dots, \mathbf{g}_{iK}]$ be the ZF beamforming matrix corresponding to the channel matrix from BS i formed by row vectors $\mathbf{h}_{i1}^H, \dots, \mathbf{h}_{iK}^H$, $\mathbf{u}_i = [\sqrt{p_{i1}} s_{i1}, \dots, \sqrt{p_{iK}} s_{iK}]^T$ is the vector of symbols where $s_{ik} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ denotes the symbol transmitted by BS i to UT k with power p_{ik} . With this beamforming, the received signal at UT k is given by

$$y_k = \sum_{i=1}^B \sqrt{p_{ik}} a_{ik} s_{ik} + n_k \quad (4)$$

where $a_{ik} = \mathbf{h}_{ik}^H \mathbf{g}_{ik}$ denotes the overall channel from BS i to UT k and coincides with the definition in (2) for each i . From Lemma 1 and the independence of $\{a_{ik}\}$ over i , we remark that the original $B \times K$ MISO interference channel (1) is decoupled into K parallel $B \times 1$ MISO channels.

Lemma 2: Let $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$ denote a MIMO slow fading channel with standard complex Gaussian noise \mathbf{z} . If we have

$$\Pr \{ \|\mathbf{H}\|_{\text{F}}^2 < \epsilon \} \doteq \epsilon^d \quad (5)$$

then the maximum diversity order of the channel is d .

Theorem 1: With distributed ZF beamforming, the diversity order of each UT is $B(M - K + 1)$.

Proof: Appendix A. ■

Assuming that each UT k perfectly knows the channel state $\mathbf{a}^k = (a_{1k}, \dots, a_{Bk})$, it decodes the space-time code and achieves the following rate

$$R_k = \log \left(1 + \sum_{i=1}^B |a_{ik}|^2 p_{ik} \right). \quad (6)$$

The capacity region of the K parallel MISO channels (4) for a *fixed* set of power $\mathbf{p}^k = (p_{1k}, \dots, p_{Bk})$ and channel state $\mathbf{a}^k = (a_{1k}, \dots, a_{Bk})$ for all k is given by

$$\mathcal{R}(\mathbf{a}; \mathbf{p}) = \left\{ \mathbf{R} \in \mathbb{R}_+^K : R_k \leq \log \left(1 + \sum_{i=1}^B |a_{ik}|^2 p_{ik} \right) \quad \forall k \right\} \quad (7)$$

where we let $\mathbf{a} = \{\mathbf{a}^k\}$, $\mathbf{p} = \{\mathbf{p}^k\}$ for a notation simplicity. The above region is clearly convex (rectangular for $K = 2$). The capacity region of the K parallel MISO channels (4) under the individual BS power constraints $\mathbf{P} = (P_1, \dots, P_B)$ for a *fixed* channel state \mathbf{a} is given by

$$\mathcal{C}(\mathbf{a}; \mathbf{P}) = \bigcup_{\mathbf{p} \in \mathcal{F}} \mathcal{R}(\mathbf{a}; \mathbf{p}) \quad (8)$$

where \mathcal{P} denotes a power allocation policy $\mathbf{a} \mapsto \mathbf{p}$ that maps the channel state \mathbf{a} into the power vector \mathbf{p} with component $\mathcal{P}_{ik}(\mathbf{a}) = p_{ik}$, \mathcal{F} denotes the feasibility set satisfying $\sum_{k=1}^K \mathcal{P}_{ik}(\mathbf{a}) \leq P_i, \forall i$ for any channel realization \mathbf{a} . The capacity region $\mathcal{C}(\mathbf{a}; \mathbf{P})$ is convex and its boundary can be explicitly characterized by solving the weighted sum rate maximization as specified in subsection III-A.

III. POWER ALLOCATION MINIMIZING OUTAGE PROBABILITY

In the previous section, we considered the distributed processing at each BS assuming that the power allocation is already done at the CS and each BS is informed about the resulting power partition. Here, we address the power allocation problem solved at the CS. In particular, we are interested in a robust power allocation strategy which requires only statistical CSIT. To this end, we consider a relevant scenario where the system imposes a set of target rate to each UT depending on its application. This is of typical of the current standards such as WiMax [20] and LTE [19]. One of the important goals in such a situation is to minimize the outage probability simultaneously for all UTs. In order to formalize the problem, we let γ_k denote the target rate of UT k and form the target rate tuple $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_K)$. For a given $\boldsymbol{\gamma}$, we define the outage probability as the average probability that $\boldsymbol{\gamma}$ is not fulfilled by all K UTs simultaneously, namely

$$\begin{aligned} P_{\text{out}}(\boldsymbol{\gamma}) &\triangleq 1 - \mathbb{E}_{\mathbf{a}}[\mathbf{1}\{\boldsymbol{\gamma} \in \mathcal{C}(\mathbf{a}, \mathbf{P})\}] \\ &= 1 - \Pr(\boldsymbol{\gamma} \in \mathcal{C}(\mathbf{a}, \mathbf{P})). \end{aligned} \quad (9)$$

This section provides the power allocation policies minimizing the outage probability defined above under perfect and statistical CSIT.

A. Perfect CSI at CS

We start with a special case where the CS has perfect CSIT. In this case, we are particularly interested in the power allocation policy that provides the rate tuple proportional to the target rate tuple (rate-balancing). As seen shortly, this policy equalizes the individual outage probability of all UTs and thus provides the strict fairness among UTs. The latter is one of the most desired properties. Our objective is find the set of $\{p_{ik}\}$ satisfying

$$\frac{R_k(\mathbf{p})}{R_1(\mathbf{p})} = \frac{\gamma_k}{\gamma_1} \triangleq \alpha_k, \quad k = 2, \dots, K \quad (10)$$

where we defined $\frac{\gamma_k}{\gamma_1} = \alpha_k$ and $\alpha_1 = 1$. More precisely, the power allocation is a solution of [22]

$$\min_{\sum_k \theta_k = 1} \max_{\mathbf{R} \in \mathcal{C}(\mathbf{a}, \mathbf{P})} \sum_{k=1}^K \theta_k \frac{R_k}{\alpha_k}. \quad (11)$$

Notice that the inner problem is the weighted sum rate maximization for a fixed weight set, while the outer problem with respect to $\theta_2, \dots, \theta_K$ is a convex problem for which a $(K - 1)$ -dimensional subgradient method can be suitably applied [23]. First, we describe a numerical method to solve the inner problem, given by

$$\max_{\sum_{k=1}^K p_{ik} \leq P_i, \forall i} \sum_{k=1}^K w_k \log \left(1 + \sum_{i=1}^B |a_{ik}|^2 p_{ik} \right) \quad (12)$$

where $w_k = \theta_k / \alpha_k$ denotes a non-negative weight for all k that we assume given for the time being. Since the above problem is convex (the objective function is concave and the constraints are linear), it is necessary and sufficient to solve the KKT condition [24], given by

$$\frac{w_k |a_{ik}|^2}{1 + \sum_{j=1}^B |a_{jk}|^2 p_{jk}} = \frac{1}{\mu_i}, \quad k = 1, \dots, K, \quad i = 1, \dots, B \quad (13)$$

where μ_i denotes the Lagrangian variable to be determined such that $\sum_k p_{ik} \leq P_i$. Unfortunately solving (13) directly seems intractable. Nevertheless, the following waterfilling approach inspired by the iterative multiuser waterfilling for the MIMO multiple access channel [25] solves the KKT conditions iteratively.

Algorithm A1 : Iterative waterfilling algorithm for B -BS weighted sum rate maximization

1) Initialize $\mathbf{p}_i^{(0)} = \mathbf{0}$ for $i = 1, \dots, B$ and $c_{ik}^{(0)} = 0$ for all i, k .

2) At each iteration n

For $i = 1, \dots, B$

- Compute $c_{ik}^{(n)} = \sum_{j \neq i} |a_{jk}|^2 p_{jk}^{(n)}$
- Waterfilling step : let $\mathbf{p}_i^{(n)}$ be

$$p_{ik}^{(n)} = \left[w_k \mu_i - \frac{1 + c_{ik}^{(n)}}{|a_{ik}|^2} \right]_+, \quad \forall k \quad (14)$$

where μ_i is determined such that $\sum_{k=1}^K p_{ik}^{(n)} = P_i$.

End

3) Continue until convergence

We have the following remarks on the proposed algorithm.

Remark 3.1: Algorithm A1 is a generalization of the classical waterfilling algorithm for the K parallel channels under the total power constraint to the case with B transmitters with individual power constraints. Indeed, for a single-BS case ($B = 1$), the objective function reduces to the weighted sum rate of the K parallel non-interfered channels given by

$$\max_{\mathbf{q} \geq \mathbf{0}, \sum_k q_k \leq P} \sum_{k=1}^K w_k \log (1 + |a_k|^2 q_k). \quad (15)$$

Remark 3.2: The set of powers in (14) correspond to the solution to the new objective function

$$\mathbf{p}_i^{(n)} = \arg \max_{\mathbf{q} \geq \mathbf{0}, \sum_k q_k \leq P_i} \sum_{k=1}^K w_k \log \left(1 + \frac{|a_{ik}|^2 q_k}{1 + c_{ik}^{(n)}} \right). \quad (16)$$

It can be easily seen that this new problem and the original problem (12) yield the same KKT conditions (13) for $c_{ik}^{(n)} = \sum_{j \neq i} |a_{jk}|^2 p_{jk}^{(n)}$. In other words, the solution (14) of BS i corresponds to treating $\{c_{ik}\}_{k=1}^K$, function of the powers $\{\mathbf{p}_j\}_{j \neq i}$ of the other BSs, as additional noise (constant), i.e., as if they did not depend on \mathbf{p}_i . Under the individual power constraints for each BS, the sequential iteration over different BSs probably converges (see the convergence proof below).

By a rather straightforward extension of [25, Theorem 1, Theorem 2], we have the following convergence result.

Theorem 2: Algorithm A1 converges to the optimal solution of the weighted sum rate maximization (12).

Proof: The proof follows in the footsteps of the proofs of [25, Theorem 1, Theorem 2]. At each iteration, the proposed iterative algorithm finds the solution to the weighted sum rate maximization of the single-BS parallel channels (15) for each BS while treating the powers of all the other BSs as additional noise. Comparing the objective of the single-BS parallel channels (15) and that of the multi-BS parallel channels (16), they differ only from the constant $c_{ik}(\{p_{jk}\}_{j \neq i})$. Hence, the objective is nondecreasing at each waterfilling step and further converges to a limit. The powers $\mathbf{p}_1, \dots, \mathbf{p}_B$ of B BSs also converge to the limit $\mathbf{p}_1^*, \dots, \mathbf{p}_B^*$. This is because the solution to the single-BS parallel channels is unique. ■

The outer problem consists of minimizing the solution of the inner problem with respect to $\theta_2, \dots, \theta_K$. This can be done by the subgradient method. We summarize the overall algorithm as follows.

Rate balancing algorithm

- 1) Initialize $\theta_k^{(0)} \in [0, 1]$ for $k > 1$.
- 2) At iteration n , compute

$$\mathbf{R}^{(n)} = \arg \max_{\mathbf{R} \in \mathcal{C}(\mathbf{a}, \mathbf{P})} \sum_{k=1}^2 \theta_k^{(n)} \frac{R_k}{\alpha_k} \quad (17)$$

via waterfilling approach.

Compute a subgradient of user k

$$\Delta_k^{(n)} = \frac{R_k^{(n)}}{\alpha_k} - R_1^{(n)}.$$

- 3) Update the weights by subgradient :

$$\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} - s_n \boldsymbol{\Delta}^{(n)}$$

where s_n denotes the step size at iteration n that can be chosen according to a decreasing rule.

The overall algorithm, the inner problem solved by algorithm A1 and the outer problem solved by the subgradient method, implements the rate-balancing by allocating the rates proportional to the target rate tuple. Hereafter, we let $R_k^*(\mathbf{a})$ denote the rate of user k allocated by the rate-balancing algorithm in channel state \mathbf{a} . Notice that the achievable rate (6) is a deterministic function of $\mathcal{P}^*(\mathbf{a})$. Then, we have the following result.

Theorem 3: The rate-balancing power allocation minimizes the outage probability. Moreover, it equalizes the individual outage probability of all K UT by letting $\Pr(\gamma_1 < R_1^*(\mathbf{a})) = \dots = \Pr(\gamma_K < R_K^*(\mathbf{a}))$.

Proof: Appendix B. ■

B. Statistical CSI at CS

We consider a more realistic case when the CS has only statistical knowledge of the equivalent channels \mathbf{a} . Formally, we define the power allocation policy $\mathcal{P}^{\text{stat}} : \boldsymbol{\sigma} \mapsto \mathbf{p}$ as a function mapping the variances of channel coefficients $\boldsymbol{\sigma} = (\sigma_{11}, \dots, \sigma_{BK})$ into the power vector \mathbf{p} with component $\mathcal{P}_{ik}^{\text{stat}}(\boldsymbol{\sigma}) = p_{ik}$. Since the equivalent channel coefficients $\{a_{ik}\}$ after ZF beamforming are correlated over different k for each BS i , the individual outage events are dependent. Nevertheless, this dependency can be made arbitrarily small by a simple interleaving over frequency bands for example. Under this assumption, we approximate the outage probability for a fixed power allocation \mathbf{p} as

$$P_{\text{app-out}}(\boldsymbol{\gamma}, \mathbf{p}) = 1 - \prod_{k=1}^K \Pr(\Delta_k(\mathbf{p}^k) > 2^{\gamma_k} - 1) \quad (18)$$

where we let $\Delta_k(\mathbf{p}^k) = \sum_{i=1}^B |a_{ik}|^2 p_{ik}$. First, we remark that for a fixed power $\mathbf{p}^k = (p_{1k}, \dots, p_{Bk})$ of UT k , $\Delta_k(\mathbf{p}^k)$ is a Hermitian quadratic form of complex Gaussian random variables given by

$$\begin{aligned} \Delta_k(\mathbf{p}^k) &= (a_{1k}^*, \dots, a_{Bk}^*) \begin{pmatrix} p_{1k} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & p_{Bk} \end{pmatrix} \begin{pmatrix} a_{1k} \\ \vdots \\ a_{Bk} \end{pmatrix} \\ &= (\mathbf{w}_{1k}^H \dots \mathbf{w}_{Bk}^H) \begin{pmatrix} p_{1k} \mathbf{I} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & p_{Bk} \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{w}_{1k} \\ \vdots \\ \mathbf{w}_{Bk} \end{pmatrix} \end{aligned}$$

where the second equality follows by replacing a chi-square random variable $|a_{ik}|^2 \sim \chi_{2(M-K+1)}^2$ with $\|\mathbf{w}_{ik}\|^2$ where $\mathbf{w}_{ik} \sim \mathcal{N}_c(\mathbf{0}, \frac{\sigma_{ik}}{M-K+1} \mathbf{I}_{M-K+1})$. The individual outage probability that UT k cannot support γ_k for a fixed \mathbf{p}^k is

$$\Pr(\Delta_k(\mathbf{p}^k) \leq c_k) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{e^{sc_k} \Phi_{\Delta_k(\mathbf{p}^k)}(s)}{s} ds \quad (19)$$

where we let $c_k = 2^{\gamma_k} - 1$ denote the target SNR of UT k , and the Laplace transform of $\Delta_k(\mathbf{p}^k)$ is given by

$$\Phi_{\Delta_k(\mathbf{p}^k)}(s) = \prod_{i=1}^B \frac{1}{(1 + s \frac{p_{ik}\sigma_{ik}}{M-K+1})^{M-K+1}}. \quad (20)$$

From (20), we see immediately that each UT achieves a diversity gain of $B(M - K + 1)$, which agrees well with Theorem 1. The widely used upper bound is the Chernoff bound for fixed powers \mathbf{p}^k given by

$$\Pr(\Delta_k(\mathbf{p}^k) \leq c_k) \leq \min_{\lambda \geq 0} e^{\lambda c_k} \Phi_{\Delta_k(\mathbf{p}^k)}(\lambda) \triangleq \bar{F}(c_k, \mathbf{p}^k). \quad (21)$$

Using the last expression of the Chernoff upper bound for each k , the approximated outage probability for a fixed \mathbf{p} is upper-bounded by

$$P_{\text{app-out}}(\gamma, \mathbf{p}) \leq 1 - \prod_{k=1}^K (1 - \bar{F}(c_k, \mathbf{p}^k)) \quad (22)$$

Since the power optimization based on the exact outage probability is not amenable, we search the power allocation that minimizes the Chernoff upper bound, equivalently solves

$$\begin{aligned} & \text{maximize} && f(\{\lambda_k\}, \{p_{ik}\}) \triangleq \prod_{j=1}^K (1 - h_k(\lambda_k, \mathbf{p}^k)) \\ & \text{subject to} && \sum_{k=1}^K p_{ik} \leq P_i, \quad i = 1, \dots, B \\ & && \lambda_k \geq 0, \quad k = 1, \dots, K \\ & && p_{ik} \geq 0, \quad i = 1, \dots, B, k = 1, \dots, K \end{aligned} \quad (23)$$

where we define

$$h_k(\lambda_k, \mathbf{p}^k) = \frac{e^{\lambda_k c_k}}{\prod_{i=1}^B (1 + \beta_{ik} \lambda_k p_{ik})^{M-K+1}} \quad (24)$$

where $\beta_{ik} = \frac{\sigma_{ik}}{M-K+1}$. If $h_k \geq 1$ for some k , the objective becomes null regardless of the power allocation. In this case a reasonable choice is to let $p_{ik} = 0, \forall i$ for such k and equally allocate the power to $\{p_{ik'}\}$ for $k' \neq k$. In the following, by focusing on the case $h_k < 1$ for all k , we provide an efficient numerical method to solve the problem (23). We remark first that the maximization of f with respect to $\{\lambda_k\}$ can be decoupled into the minimization of h_k over λ_k for each k , where h_k is convex in λ_k . Moreover, since f is concave in $\{p_{ik}\}$, the overall problem is convex.

Minimization of h_k over λ_k It can be easily verified that h_k is monotonically decreasing in λ_k . The optimal λ_k for a fixed set of powers is the solution of

$$\frac{c_k}{M - K + 1} = \sum_{i=1}^B \frac{\beta_{ik} p_{ik}}{1 + \lambda_k \beta_{ik} p_{ik}} \quad (25)$$

which is a polynomial of degree B . For $B = 2$, the solution is given in a closed form.

$$\lambda_k = \begin{cases} \frac{M-K+1}{c_k} - \frac{1}{\beta_{1k}p_{1k} + \beta_{2k}p_{2k}}, & \text{if } \sum_{j=1}^2 \beta_{jk}p_{jk} = 0 \\ \frac{-\left(\beta_{1k}p_{1k} + \beta_{2k}p_{2k} - \frac{2(M-1)\beta_{1k}p_{1k}\beta_{2k}p_{2k}}{c_k}\right) + \sqrt{\left(\beta_{1k}p_{1k} - \beta_{2k}p_{2k}\right)^2 + \left(\frac{2(M-1)\beta_{1k}p_{1k}\beta_{2k}p_{2k}}{c_k}\right)^2}}{2\beta_{1k}p_{1k}\beta_{2k}p_{2k}}, & \text{otherwise} \end{cases} \quad (26)$$

Maximization of f over p_{ik} Since f is concave in $\{p_{ik}\}$, we form the Lagrangian function by introducing B Lagrangian multipliers $\{\mu_i\}$ each of which is associated to the power constraint of BS i . By arranging the term common for all k , we obtain the KKT conditions for $k = 1, \dots, K$

$$\frac{h_k(\mathbf{p}^k)}{1 - h_k(\mathbf{p}^k)} \frac{\beta_{ik}\lambda_k}{1 + p_{ik}\beta_{ik}\lambda_k} = \mu_i. \quad (27)$$

When treating $p_{1k}, \dots, p_{i-1,k}, p_{i+1,k}, \dots, p_{B,k}$ fixed, the LHS of (27), denoted by ϕ_{ik} , is a strictly positive and monotonically decreasing function of p_{ik} (since we exclude the case $\lambda_k = 0$). It remains to determine μ_i such that the power constraint of BS i is satisfied, i.e., $p_{i1} + \dots + p_{iK} = P_i$. When treating the powers $\{\mathbf{p}_j\}_{j \neq i}$ of the other BSs $j \neq i$ fixed, the powers \mathbf{p}_i of BS i can be found by a simple line search of μ_i .

The following summarizes our proposed iterative algorithm to minimize the Chernoff upper bound, equivalently solve (23).

Algorithm A2 : iterative algorithm for the Chernoff upper bound minimization

- 1) Initialize $\mathbf{p}^{(0)}$
- 2) At iteration n
 - For $i = 1, \dots, B$
 - Update $\lambda^{(n)}$ by solving the polynomial (25)
 - Find the new power vector $\mathbf{p}_i^{(n)}$ of BS i by line search
 - End
- 3) Continue until converge f

Although we are unable to provide a formal proof, we conjecture that Algorithm A2 converges to its optimal solution. At each iteration, $\lambda_k^{(n)}$ is determined as a unique solution for all k and a fixed set of powers. Regarding the power iteration, since the objective (27) is concave in p_{ik} when fixing all other powers, a sequential update of the powers $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_B, \mathbf{p}_1, \dots$ shall converge under individual BS power constraints by a similar argument as the proof of Theorem 2.

IV. EFFECT OF USER SCHEDULING

In this section we address the relevant case when the number of UTs is larger than the number of transmit antennas $K > M$. In order for each BS to apply the ZF beamforming in a distributed fashion,

a set of $\tilde{K} \leq M$ UTs shall be selected beforehand. We assume that the user selection (scheduling) is handled by the CS together with the power allocation. In particular, we focus on a user selection method which achieves a high diversity order while limiting the amount of the side information necessary at the CS and the BSs. In the following, we present our proposed user selection scheme as well as the analysis on its achievable diversity gain.

A. Distributed Diversity Scheduling (DDS)

Let \mathcal{S}, \mathcal{U} denote the set of all K users, the \tilde{K} selected users, with $|\mathcal{S}| = K$, $|\mathcal{U}| = \tilde{K}$, respectively. Let us also define $\mathcal{Q}(\tilde{K})$ as the set of all possible user selections, i.e., $\mathcal{Q}(\tilde{K}) = \{\mathcal{U} \mid \mathcal{U} \subseteq \mathcal{S}, |\mathcal{U}| = \tilde{K}\}$ for $\tilde{K} \leq M$.² Then, the equivalent channel from the BSs to the selected users is

$$y_k = \mathbf{a}^k \mathbf{u}^k + z_k, \quad k \in \mathcal{U}, \quad (28)$$

which is a MISO channel with $\mathbf{a}^k = [a_{1k} \cdots a_{Bk}]$ and $\mathbf{u}^k = [\sqrt{p_{1k}} s_{1k} \cdots \sqrt{p_{Bk}} s_{Bk}]^T$. For convenience, we only consider the diversity order of the worst user and refer it as the diversity of the system hereafter. Since the diversity order of a given channel depends solely on the Euclidean norm of the channel matrix, as shown in lemma 2, the following user selection scheme maximizes the diversity of the system

$$\mathcal{U}^* = \arg_{\mathcal{U}} \max_{\mathcal{U} \in \mathcal{Q}} \min_{k \in \mathcal{U}} \|\mathbf{a}^k\|^2. \quad (29)$$

Unfortunately, this scheduling scheme has two major drawbacks: 1) perfect knowledge at the CS on $\{\mathbf{a}_k\}$, crucial for the scheduling, is hardly implementable as aforementioned, and 2) the maximization over all $|\mathcal{Q}(\tilde{K})| = \binom{K}{\tilde{K}}$ possible sets \mathcal{U} grows in polynomial time with K .

To overcome the first drawback, we use the following selection scheme

$$\mathcal{U}_d = \arg_{\mathcal{U}} \max_{i=1 \dots B} \max_{\mathcal{U} \in \mathcal{Q}} \min_{k \in \mathcal{U}} |a_{ik}|^2. \quad (30)$$

That is, BS i finds out the set \mathcal{U} that maximizes $\min_{k \in \mathcal{U}} |a_{ik}|^2$ and sends both the index of the set and the corresponding maximum value to the CS. Upon the reception of B values and the corresponding sets from the B BS, the CS makes a decision by selecting the largest one. Therefore, only partial channel state information is communicated in the BS-CS link. To address the second drawback, we narrow down the choices of \mathcal{U} to the following $\kappa = K/\tilde{K}$ possibilities³

$$\mathcal{P}_{\mathcal{S}} = \{\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_{\kappa}\}, \quad \bigcup_i \mathcal{U}_i = \mathcal{S}, \quad \mathcal{U}_i \cap \mathcal{U}_j = \emptyset, \quad \forall i \neq j, \quad |\mathcal{U}_i| = \tilde{K}, \quad \forall i.$$

²For convenience of notation, we will drop the argument \tilde{K} whenever confusion is unlikely.

³Here, we assume that K/\tilde{K} is integer for simplicity of demonstration. However, it will be shown that same conclusion holds otherwise.

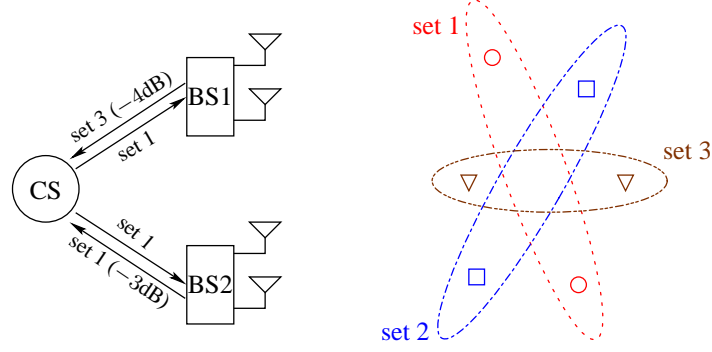


Fig. 2. An example scenario of user scheduling with two BSs and six UTs.

In other words, \mathcal{P}_S is partition of the set of all users \mathcal{S} . Furthermore, it is assumed that the partition \mathcal{P}_S is fixed by the CS and known to all BSs. Hence, the proposed scheduling scheme selects the following set of users

$$\mathcal{U}_d = \arg \max_{\mathcal{U}} \max_{i=1 \dots B} \min_{k \in \mathcal{U}} |a_{ik}|^2. \quad (31)$$

To summarize, the scheduling scheme works as follows

- i) The CS fixes a partition \mathcal{P}_S and informs it to all BSs.
- ii) BS i finds $\max_{\mathcal{U} \in \mathcal{P}_S} \min_{k \in \mathcal{U}} |a_{ik}|^2$, and sends this value and the index of the maximizing set \mathcal{U} to the CS.
- iii) The CS chooses the largest value and broadcasts the index of the winner set \mathcal{U}_d as defined in (31).
- iv) All the BSs serve simultaneously the UTs in \mathcal{U}_d .

An example of two BSs and six UTs is shown in Fig. 2. In this example, in order to serve two UTs simultaneously, a partition of three sets is fixed by the CS. With local CSI, each BS compares the coefficients $\min_{k \in \mathcal{U}} |a_{ik}|^2$ for all three sets \mathcal{U} , finds out the largest one, and sends the corresponding “index(value)” pair to the CS. The CS compares the values and broadcasts the index of the winning set (set 1 in this example).

B. Diversity gain analysis

In this subsection, we analyze the diversity gain achieved by our proposed DDS and compare it with the upper bound. The result is summarized in the following Theorem.

Theorem 4: Let K , B , and M denote the number of UTs, number of BSs, and number of antennas

per BS. By serving \tilde{K} UTs simultaneously, the following diversity gain is achievable with DDS

$$d_{\mathcal{U}_d}(\tilde{K}) \geq B \frac{K}{\tilde{K}} (M - \tilde{K} + 1), \quad (32)$$

for $K = n\tilde{K}$ with an integer n . Furthermore, the optimal diversity gain achieved by (29) is upper-bounded by

$$d_{\mathcal{U}^*}(\tilde{K}) \leq B(K - \tilde{K} + 1) (M - \tilde{K} + 1). \quad (33)$$

Proof: Appendix C. ■

Remark 4.1: For a fixed \tilde{K} , the diversity gain of the proposed scheduling scheme grows as $O(BKM)$, i.e., the optimal diversity scaling with B , K , and M . In this sense, the proposed scheme is order optimal in terms of diversity gain. Also note that the lower bound (32) and the upper bound (33) coincide in some specific settings. First, $d_{\mathcal{U}_d}(1) = d_{\mathcal{U}^*}(1) = BKM$, $\forall B, K, M$. That is, the proposed scheme is diversity optimal if only one user is served in the system. Then, for $K \leq M$, we have $d_{\mathcal{U}_d}(K) = d_{\mathcal{U}^*}(K) = B(M - K + 1)$. This corresponds to the case where all users in the system are served simultaneously.

Remark 4.2: Interestingly, exactly the same diversity order is achieved for \mathcal{U}^* if we set $\mathcal{Q} = \mathcal{P}_s$. To see this, let us rewrite

$$\min_{k \in \mathcal{U}^*} \|\mathbf{a}^k\|^2 = \max_{\mathcal{U} \in \mathcal{P}_s} \min_{k \in \mathcal{U}} \|\mathbf{a}^k\|^2 \quad (34)$$

$$\leq \max_{\mathcal{U} \in \mathcal{P}_s} \|\mathbf{a}^k\|^2, \quad \forall k \in \mathcal{U} \quad (35)$$

and that $\max_{\mathcal{U} \in \mathcal{P}_s} \|\mathbf{a}^k\|^2$ is of diversity $B \frac{K}{\tilde{K}} (M + 1 - \tilde{K})$.

Remark 4.3: When \tilde{K} does not divide K , we consider only $\left\lfloor \frac{K}{\tilde{K}} \right\rfloor \tilde{K}$ out of K users. Since \tilde{K} divides $\left\lfloor \frac{K}{\tilde{K}} \right\rfloor \tilde{K}$, the following diversity gain can be achieved

$$d_{\mathcal{U}_d}(\tilde{K}) = B \left\lfloor \frac{K}{\tilde{K}} \right\rfloor (M - \tilde{K} + 1) \quad (36)$$

with the proposed DDS.

V. NUMERICAL EXAMPLES

This section provides some numerical examples to verify the behavior of our proposed distributed ZF beamforming scheme in a simple network MIMO configuration with $B = 2$ cooperative BSs. We assume the same power constraint at both BSs $P_1 = P_2$ and let P denote the SNR.

Fig. 5 shows the outage probability performance versus SNR for $K = 2$ and $M = 2, 4$. The target rate is fixed to $(\gamma_1, \gamma_2) = (3, 1)$ bit/channel use, and we let $\sigma_{ik} = 1$ for all i, k . We compare the different

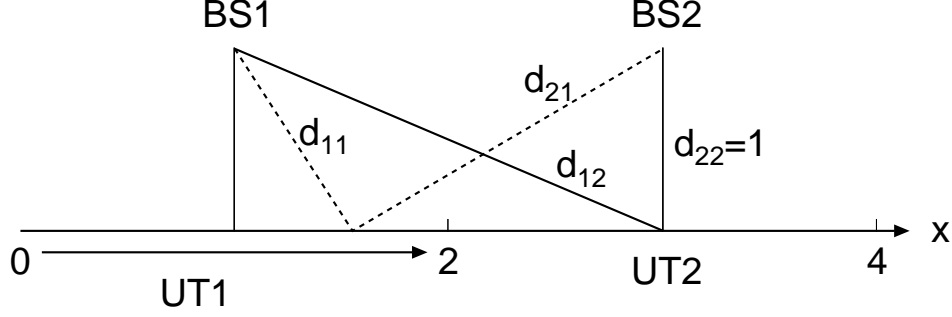


Fig. 3. One dimensional configuration.

power allocation strategies, algorithm A1 with perfect CSIT, algorithm A2 with statistical CSIT, and equal power allocation ($p_{i1} = p_{i2} = P/2$ for $i = 1, 2$). For the sake of comparison, we also consider the case without network MIMO (no message sharing) where each BS sends a message to its corresponding UT in a distributed fashion. In order to make the comparison fair in terms of complexity, we let each BS i send the symbol s_i by ZF beamforming, i.e. $\mathbf{x}_i = \sqrt{P}\mathbf{g}_i s_i$ where \mathbf{g}_i is a unit-norm vector orthogonal to \mathbf{h}_{ik} for $k \neq i$. From Lemma 1, such a system offers a diversity order of $M - K + 1$ for each UT. As expected from Theorem 1, we observe that our BS cooperation schemes enables to achieve a diversity gain of $2(M - K + 1)$, i.e. 2, 6 with $M = 2, 4$, respectively. These gains are twice as large as the case without network MIMO. Moreover, the proposed algorithms provide a significant power gain compared to equal power allocation.

In Fig. 6, we plot the individual outage probability such that each UT k cannot support its target rate γ_k under the same setting as Fig. 5 for $M = 2$. With perfect CSIT, our proposed waterfilling allocation A1 guarantees the identical outage probability for both UTs by offering the strict fairness. This agrees well with the second part of Theorem 3. Under statistical CSIT, algorithm A2 provides a better outage probability to UT 1 but keeps the gap between two UTs smaller than the equal power allocation.

In order to evaluate the impact of asymmetric path loss on the outage performance, we consider a simple 1-D configuration illustrated in Fig. 3 where UT 2 is located at $x = 3$ and UT 1 moves from $x = 0$ to $x = 2$. Assuming that BS 1, 2 is at $x = 1, 3$, respectively, we vary $d_{11} = \sqrt{1 + (1 - x)^2}$, $d_{12} = \sqrt{1 + (3 - x)^2}$ while we fix the position of UT2 by letting $d_{12} = \sqrt{5}$, $d_{22} = 1$. By taking into account the path loss $\sigma_{ik} = d_{ik}^{-3}$, we plot the outage probability as a function of the position x of UT 1 in Fig. 7. We consider $M = 4$, SNR $P = 10$ dB and fix the target rate $\gamma_1 = \gamma_2 = 1$ bit/channel use. We observe that the proposed distributed ZF scheme provides a significant gain compared to the case without network

MIMO especially as UT 1 gets closed to the cell boundary ($x = 2$). This is because the performance without network MIMO only depends on d_{11} while the distributed ZF becomes more and more beneficial as d_{21} decreases.

Finally, Fig. 8 shows the outage probability versus SNR when we have more users than the number of served users, i.e. $K \geq \tilde{K} = 2$. Considering the same setting as Fig. 5 for $M = 4$, we apply distributed diversity scheme to select a set of two users among $K = 2, 4, 6$. Once the user selection is done, any power allocation studied in Section III can be applied. However, it is non-trivial (if not impossible) to characterize the statistics of the overall channel gains in the presence of any user scheduling. Hence, we illustrate here only the performance with equal power allocation under statistical CSIT. As a matter of fact, any smarter allocation shall perform between the waterfilling allocation and the equal power allocation. As expected from Theorem 4, the diversity gain increases significantly as the number K of users in the system gets large.

VI. CONCLUSIONS

We considered the multi-cell downlink system (network MIMO) where B BSs, perfectly connected via the reliable backbone links to the CS, wish to communicate simultaneously with K UTs. As one of the realistic limitations of network MIMO, we explicitly accounted for partial CSIT, i.e. local channel knowledge at each BS and statistical channel knowledge at the CS. Under this setting, we proposed an outage-efficient strategy which builds on distributed ZF beamforming to be performed at each BS and efficient power allocation algorithms at the CS. For the case of a small number of users $K \leq M$, the proposed scheme enables each UT to achieve a diversity gain of $B(M - K + 1)$. For the case of many users $K \geq M$, we proposed distributed diversity scheduling (DDS) which can be implemented in a distributed fashion at each BS and requires only limited amount of the backbone communications. We also proved that DDS can offer the diversity gain of $B\frac{K}{\tilde{K}}(M - \tilde{K} + 1)$ and this gain scales optimally with the number of cooperative BSs as well as the number of UTs. The main finding is that limited BS cooperation can still make network MIMO attractive in the sense that a well designed scheme can offer high data rates with sufficient reliability to individual UTs. The proposed scheme can be suitably applied to any other interference networks where the transmitters can perfectly share the messages to all UTs and a master transmitter can handle the resource allocation.

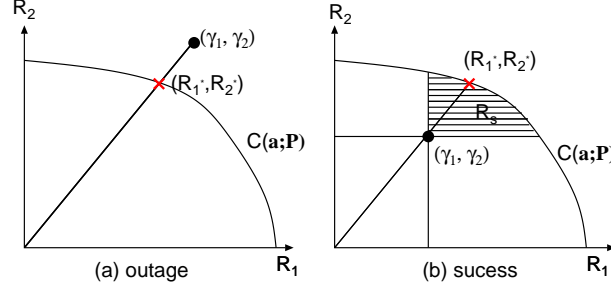


Fig. 4. Rate region $\mathcal{C}(\mathbf{a}; \mathbf{P})$ and subregion \mathcal{R}_s

APPENDIX

A. Proof of Theorem 1

The channel (4) is a MISO channel defined by $\mathbf{a}^k = (a_{1k}, \dots, a_{Bk})$ and

$$\Pr \left\{ \|\mathbf{a}^k\|^2 < \epsilon \right\} \doteq \Pr \left\{ \max_i |a_{ik}|^2 < \epsilon \right\} \quad (37)$$

$$= \prod_i \Pr \left\{ |a_{ik}|^2 < \epsilon \right\} \quad (38)$$

$$= \left(\Pr \left\{ |a_{ik}|^2 < \epsilon \right\} \right)^B \quad (39)$$

$$\doteq \epsilon^{B(M-K+1)} \quad (40)$$

From lemma 2, the maximum diversity is $B(M - K + 1)$.

B. Proof of Theorem 3

First we remark that for a given channel realization \mathbf{a} the following two cases occur:

- (a) The target rate tuple is outside the region $\boldsymbol{\gamma} \notin \mathcal{C}(\mathbf{a}, \mathbf{P})$
- (b) The target rate tuple is inside the region $\boldsymbol{\gamma} \in \mathcal{C}(\mathbf{a}, \mathbf{P})$

The above two cases are illustrated in Fig.4(a), (b) respectively for $K = 2$. For the case (a), we are in an outage event regardless of the power allocation. For the case (b), we define the subregion $\mathcal{R}_s(\mathbf{a})$

$$\mathcal{R}_s(\mathbf{a}) \triangleq \{\mathbf{R} | \mathbf{R} \in \mathcal{C}(\mathbf{a}, \mathbf{P}), R_k \geq \gamma_k, k = 1, \dots, K\} \quad (41)$$

depicted in a shadow area in Fig. 4 (b). We define \mathcal{P}_s as a class of the power allocation policies that maps \mathbf{a} into the rate tuple \mathbf{R} inside $\mathcal{R}_s(\mathbf{a})$ whenever we are in case (b). We remark that any policy belonging to \mathcal{P}_s results in a successful transmission, and thus minimizes the outage probability. Since

the proposed rate balancing scheme allocates the power (p_1^*, \dots, p_K^*) so that the rate-tuple (R_1^*, \dots, R_K^*) is proportional to γ on the boundary of $\mathcal{C}(\mathbf{a}, \mathbf{P})$ whenever $\gamma \in \mathcal{C}(\mathbf{a}, \mathbf{P})$, it belongs to the class \mathcal{P}_s . This establishes the first part.

We now prove the second part. It is immediate to see that with the rate balancing scheme, we have for any \mathbf{a}

$$\mathbf{1}(\gamma_k < R_k^*(\mathbf{a})) = \mathbf{1}(\alpha_k \gamma_1 < \alpha_k R_1^*(\mathbf{a})) = \mathbf{1}(\gamma_1 < R_1^*(\mathbf{a})), \quad k = 2, \dots, K \quad (42)$$

where the first equality follows from (10). The outage probability with the rate balancing scheme can be always written as

$$\begin{aligned} P_{\text{out}}^{\text{balance}}(\gamma) &= 1 - \Pr(\cap_{k=1}^K \{\gamma_k < R_k^*(\mathbf{a})\}) \\ &= 1 - \Pr(\gamma_k < R_k^*(\mathbf{a})) \Pr(\cap_{k' \neq k} \{\gamma_{k'} < R_{k'}^*(\mathbf{a})\} | \gamma_k < R_k^*(\mathbf{a})) \\ &= 1 - \Pr(\gamma_k < R_k^*(\mathbf{a})) \end{aligned}$$

where the last equality follows since the equalities (42) imply $\Pr(\cap_{k' \neq k} \gamma_{k'} < R_{k'}^*(\mathbf{a}) | \gamma_k < R_k^*(\mathbf{a})) = 1$ for any k . This completes the second part.

C. Proof of Theorem 4

In order to examine the diversity gain of the proposed scheduling scheme, we first remark

$$\min_{k \in \mathcal{U}_d} \|\mathbf{a}^k\|^2 \geq \min_{k \in \mathcal{U}_d} \max_i |a_{ik}|^2 \quad (43)$$

$$\geq \max_i \min_{k \in \mathcal{U}_d} |a_{ik}|^2 \quad (44)$$

$$= \max_{\mathcal{U} \in \mathcal{P}_s} \max_i \min_{k \in \mathcal{U}} |a_{ik}|^2 \quad (45)$$

where the (44) follows from the max-min inequality and the last equality holds since we can rewrite (31) as

$$\mathcal{U}_d = \arg \max_{\mathcal{U}} \max_{i=1 \dots B} \min_{k \in \mathcal{U}} |a_{ik}|^2. \quad (46)$$

by swapping the maximization over \mathcal{U} and that over i . To find the diversity order of the scheme, we need to look at the following near-zero behavior of the channel coefficients

$$\Pr \left\{ \min_{k \in \mathcal{U}_d} \|\mathbf{a}^k\|^2 < \epsilon \right\} \leq \Pr \left\{ \max_{\mathcal{U} \in \mathcal{P}_s} \max_i \min_{k \in \mathcal{U}} |a_{ik}|^2 < \epsilon \right\} \quad (47)$$

$$= \left(\Pr \left\{ \min_{k \in \mathcal{U}} |a_{bk}|^2 < \epsilon \right\} \right)^{B|\mathcal{P}_s|} \quad (48)$$

$$\leq \left(\sum_{k \in \mathcal{U}} \Pr \{ |a_{bk}|^2 < \epsilon \} \right)^{B|\mathcal{P}_s|} \quad (49)$$

$$= (|\mathcal{U}| \Pr \{ |a_{bk}|^2 < \epsilon \})^{B|\mathcal{P}_s|} \quad (50)$$

$$\doteq (|\mathcal{U}| \epsilon^{M+1-\tilde{K}})^{B|\mathcal{P}_s|} \quad (51)$$

$$\doteq \epsilon^{B \frac{K}{\tilde{K}} (M+1-\tilde{K})}, \quad (52)$$

where (48) follows from the fact that \mathcal{U} 's are disjoint in \mathcal{P}_s and that $\min_{k \in \mathcal{U}} |a_{ik}|^2$ are independent for different \mathcal{U} and i ; (49) is from the union bound. From lemma 2, (32) is straightforward. For the upper bound of the diversity gain of the scheduling scheme (29), let us first write

$$\max_{\mathcal{U} \in \mathcal{Q}} \min_{k \in \mathcal{U}} \|\mathbf{a}^k\|^2 \leq \max_{\mathcal{U} \in \mathcal{Q}} \|\mathbf{a}^1\|^2 \quad (53)$$

$$\leq B \max_i \max_{\mathcal{U} \in \mathcal{Q}} |a_{i1}|^2 \quad (54)$$

where the first inequality is from the fact that the worst user cannot be better than the first user; the second inequality is from $\|\mathbf{a}^1\|^2 \leq B \max_i |a_{i1}|^2$. From [26, Theorem 1], we know that the diversity gain of $|a_{i1}|^2$ is $(K - \tilde{K} + 1)(M - \tilde{K} + 1)$. Therefore, it readily follows that the diversity gain of $\max_{\mathcal{U} \in \mathcal{Q}} \min_{k \in \mathcal{U}} \|\mathbf{a}^k\|^2$ is upper-bounded by $B(K - \tilde{K} + 1)(M - \tilde{K} + 1)$.

ACKNOWLEDGMENT

This work is partially supported by the European Commission through the FP7 project WiMAGIC (www.wimagic.eu) and by the French cluster System@tic through the project POSEIDON. The authors would like to thank Jakob Hoydis for his help in generating some numerical examples.

REFERENCES

- [1] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Trans. on Wireless Communications*, vol. 13, no. 4, pp. 56–61, 2006.
- [2] S. Shamai and B. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitter end," in *Proc. IEEE Vehicular Technology Conf. VTC-Spring*, pp. 1745–1749.

- [3] O. Somekh, B. Zaidel, and S. Shamai, "Sum Rate Characterization of Joint Multiple Cell-Site Processing," *IEEE Trans. on Inform. Theory*, vol. 53, no. 12, pp. 4473–4497, Dec. 2007.
- [4] A. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. on Inform. Theory*, vol. 40, no. 6, pp. 1713–1727, June 1994.
- [5] O. Somekh, O. Simeone, Y. Bar-Ness, A. Haimovich, U. Spagnolini, and S. Shamai, "An information theoretic view of distributed antenna processing in cellular systems," *Distributed antenna systems: open architecture for future wireless communications*, 2007.
- [6] S. Jing, D. N. C. Tse, J. B. Soriaga, J. Hou, J. Smee, and R. Padovani, "Downlink Macro-Diversity in Cellular Networks," in *Proc. IEEE International Symposium on Information Theory*, 2007, pp. 1–5.
- [7] O. Somekh, O. Simeone, Y. Bar-Ness, A. Haimovich, and S. Shamai, "Distributed multi-cell zero-forcing beamforming in cellular downlink channels," in *Proc. IEEE Global Telecommunications Conference*, 2006.
- [8] M. A. Maddah-Ali, S. A. Motahari, and A. K. Khandani, "Communication over MIMO X Channels: Signaling and Performance Analysis," *Technical Report University of Waterloo E&CE 2006-27*.
- [9] G. Caire, S. A. Ramprasad, H. C. Papadopoulos, C. Pepin, and C.-E. W. Sundberg, "Multiuser MIMO Downlink with Limited Inter-Cell Cooperation: Approximate Interference Alignment in Time, Frequency, and Space," *Allerton' 2008*, 2008.
- [10] W. Choi and J. G. Andrews, "The Capacity Gain from Base Station Cooperative Scheduling in a MIMO DPC Cellular System," in *Proc. IEEE International Symposium on Information Theory*, 2006, pp. 1224–1228.
- [11] S. A. Jafar and S. Shamai, "Degrees of Freedom Region for the MIMO X Channel," *IEEE Trans. on Inform. Theory*, vol. 54, Jan. 2008.
- [12] J. Hoydis, M. Kobayashi, and M. Debbah, "On the optimal number of cooperative base stations in network mimo," December 2009, submitted to IEEE Symposium on Information Theory (ISIT) 2010, preprint available <http://arxiv.org/abs/0912.4595>.
- [13] O. Simeone, O. Somekh, H. Poor, and S. Shamai, "Downlink multicell processing with limited-backhaul capacity," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 1–10, 2009.
- [14] S. Shamai, O. Somekh, O. Simeone, A. Sanderovich, B. Zaidel, and H. Poor, "Cooperative Multi-Cell Networks: Impact of Limited-Capacity Backhaul and Inter-Users Links," *Arxiv preprint arXiv:0710.3888*, 2007.
- [15] P. Marsch and G. Fettweis, "A Framework for Optimizing the Downlink Performance of Distributed Antenna Systems under a Constrained Backhaul," in *Proceedings of the 13th European Wireless Conference (EW'07), Paris*, 2007.
- [16] T. Tamaki, K. Seong, and J. Cioffi, "Downlink MIMO Systems Using Cooperation Among Base Stations in a Slow Fading Channel," in *Proc. IEEE International Conference on Communications*, June 2007, pp. 4728–4733.
- [17] B. L. Ng, J. S. Evans, S. V. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. on Inform. Theory*, vol. 54, no. 12, pp. 5491–5499, Dec. 2008.
- [18] N. Levy and S. Shamai, "Clustered local decoding for Wyner-type cellular models," in *Proc. Information Theory and Applications Workshop*, 2009, pp. 318–322.
- [19] S. Sesia, I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice*. Wiley Publishing, 2009.
- [20] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*. Prentice Hall PTR, 2007.
- [21] E. Björnson, R. Zakhour, D. Gesbert, and B. Ottersten, "Distributed multicell and multiantenna precoding: Characterization and performance evaluation," in *Proc. IEEE Global Telecommunications Conference*, 2009.

- [22] J. Lee and N. Jindal, "Symmetric capacity of MIMO downlink channels," in *Proc. IEEE International Symposium on Information Theory*, Seattle, WA, 2006.
- [23] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA : Athena Scientific, 1999.
- [24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [25] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative Water-Filling for Gaussian Vector Multiple-Access Channels," *IEEE Trans. on Inform. Theory*, vol. 50, Jan. 2004.
- [26] J. Jalden and B. Ottersten, "On the maximal diversity order of spatial multiplexing with transmit antenna selection," *IEEE Trans. on Inform. Theory*, vol. 53, no. 11, pp. 4273–4276, Nov. 2007.

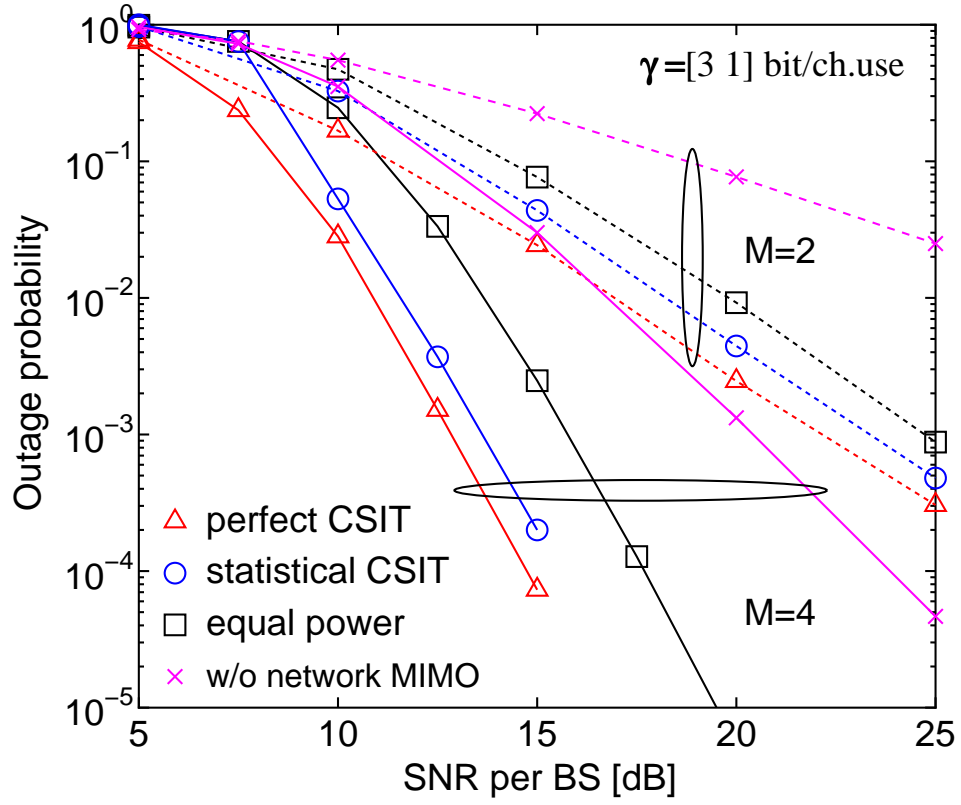


Fig. 5. Outage probability vs. SNR with $B = K = 2$ and $M = 2, 4$.

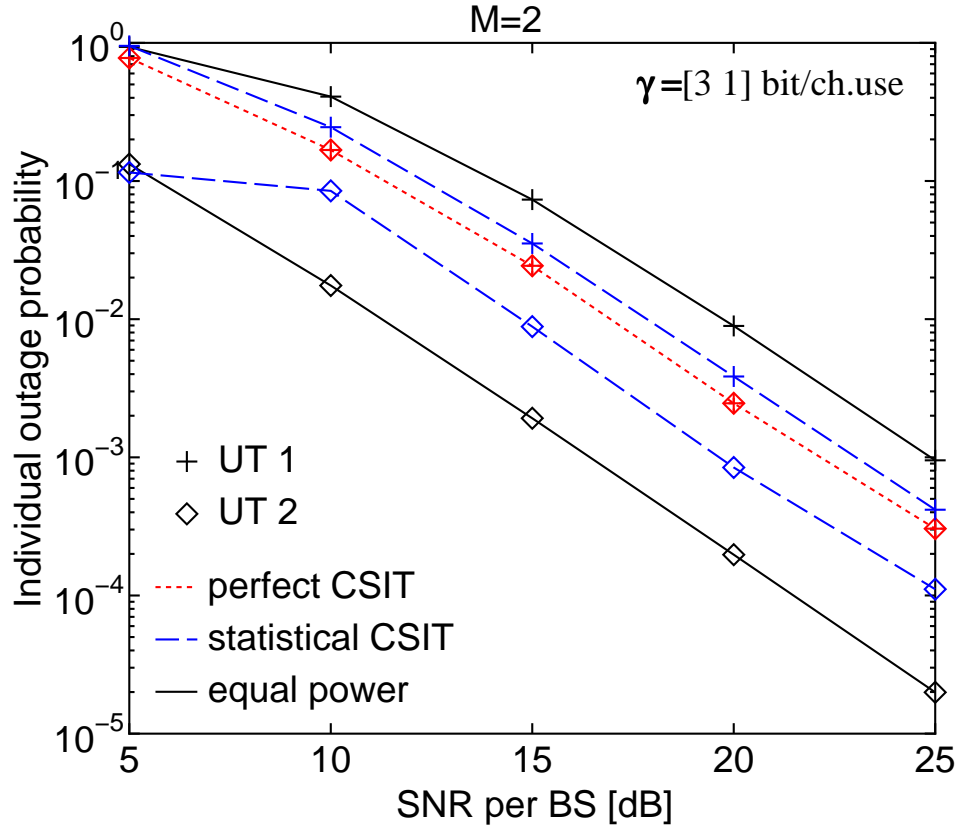


Fig. 6. Individual outage probabilities vs. SNR with $B = K = 2$ and $M = 2$.

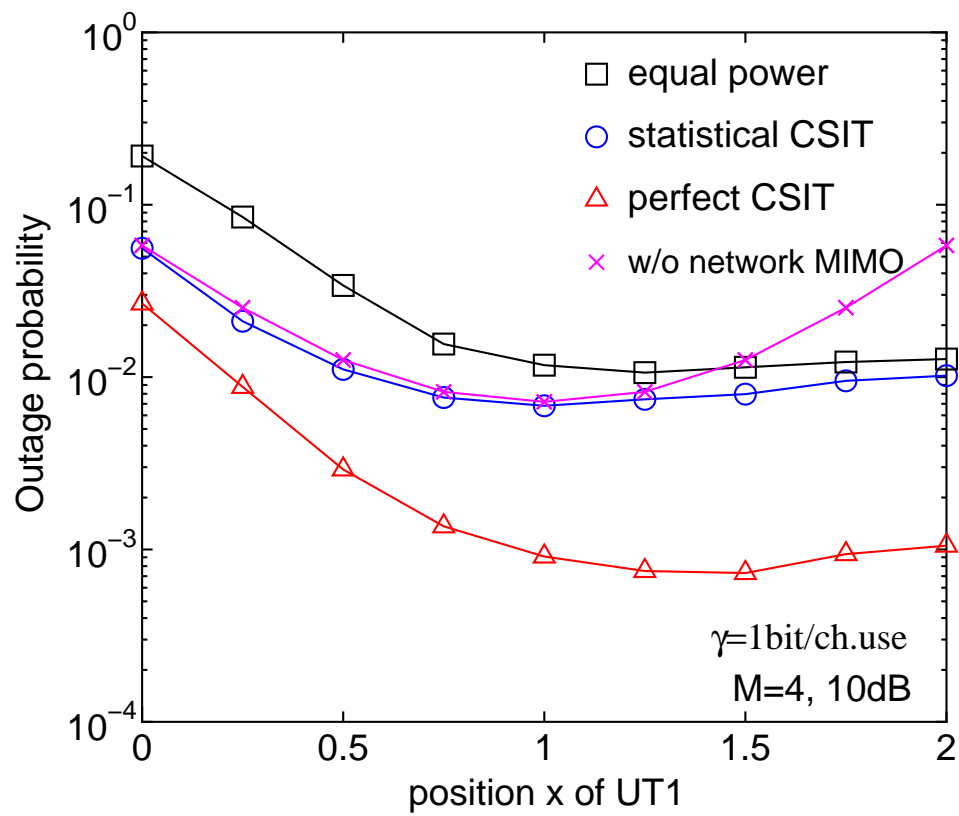


Fig. 7. Outage probability vs. location of UT1 with $B = K = 2$ and $M = 2, 4$.

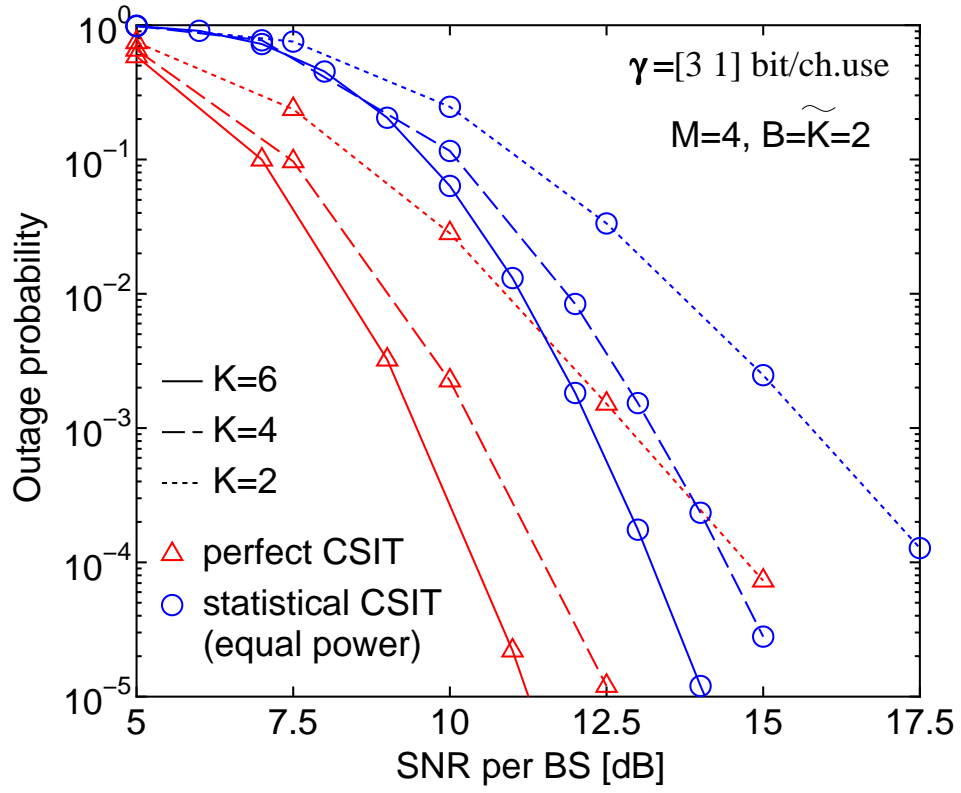


Fig. 8. Outage probability vs. SNR for many users with $B = \tilde{K} = 2$ and $M = 4$.